FIGURE 1 DOCUMENT CORPUS

DOC1: Hello, my name is Fred.
DOC2: Hello, my name is Scott.
DOC3: Scott says, "Live and let live."

FIGURE 2 DICTIONARY

1. name
2. fred
3. scott
4. Live

FIGURE 3 DENSE MATRIX - INTEGER FORMAT
(PRIOR ART)

```
1 1 0 0
1 0 1 0
0 0 1 2
```

(Space required:
16 short ints at 2 bytes each = 32 bytes)

FIGURE 4 DENSE MATRIX - FLOATING POINT NUMBER FORMAT
(PRIOR ART)

```
0.707  0.707  0.0    0.0
0.707  0.0    0.707  0.0
0.0    0.0    0.447  0.894
```

(Space required:
12 floats at 4 bytes each
= 48 bytes)

(Note: $0.707 = 1*1/(1^2 + 1^2)^{1/2}$; $0.447 = 1*1/(1^2 + 2^2)^{1/2}$; $0.894 = 2*1/(1^2 + 2^2)^{1/2}$)

FIGURE 5 SPARSE MATRIX - FLOATING POINT NUMBER FORMAT
(PRIOR ART)

```
(1  0.707) (2  0.707)
(1  0.707) (3  0.707)
(3  0.447) (4  0.894)
```

(Space required:
6 short ints & 6 floats
= 6 * 2 + 6 * 4 = 36 bytes)

FIGURE 6 SMALL SPARSE MATRIX - FLOATING POINT NUMBER FORMAT

1, 2: 0.707
1, 3: 0.707
3,4,4: 0.447

(Space required:
7 short ints & 3 floats
= 7*2 + 3*4 = 26 bytes)

(Note: $0.707 = 1/(1^2 + 1^2)^{1/2}$; $0.447 = 1/(1^2 + 2^2)^{1/2}$)

FIGURE 7 SMALL SPARSE MATRIX IN VECTOR FORM

ALLDATA = 1 2 1 3 3 4 4
STARTMARKER = 1,3,5
MULT = 0.707 0.707 0.447

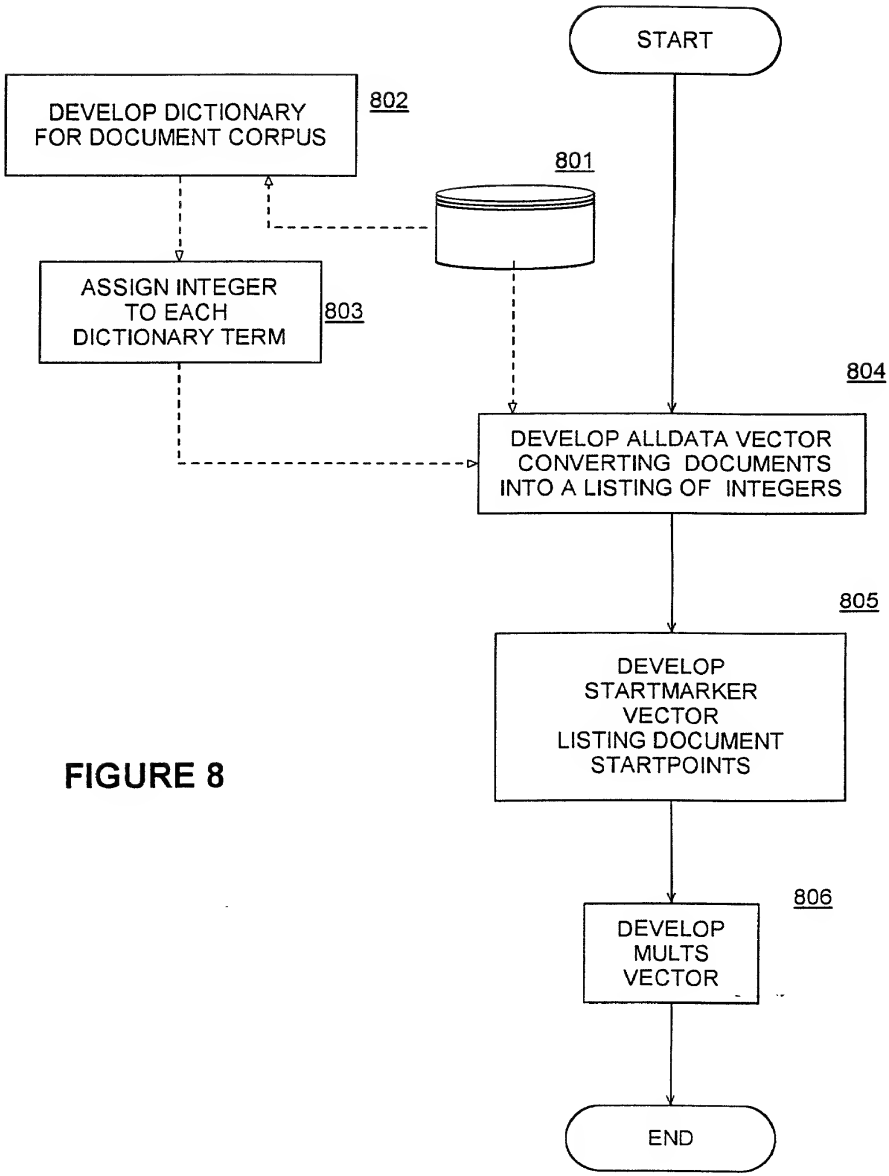

START
DEVELOP DICTIONARY FOR DOCUMENT CORPUS
802
801
ASSIGN INTEGER TO EACH DICTIONARY TERM
803
804
DEVELOP ALLDATA VECTOR CONVERTING DOCUMENTS INTO A LISTING OF INTEGERS
805
DEVELOP STARTMARKER VECTOR LISTING DOCUMENT STARTPOINTS
806
DEVELOP MULTS VECTOR
END

FIGURE 8

FIGURE 9

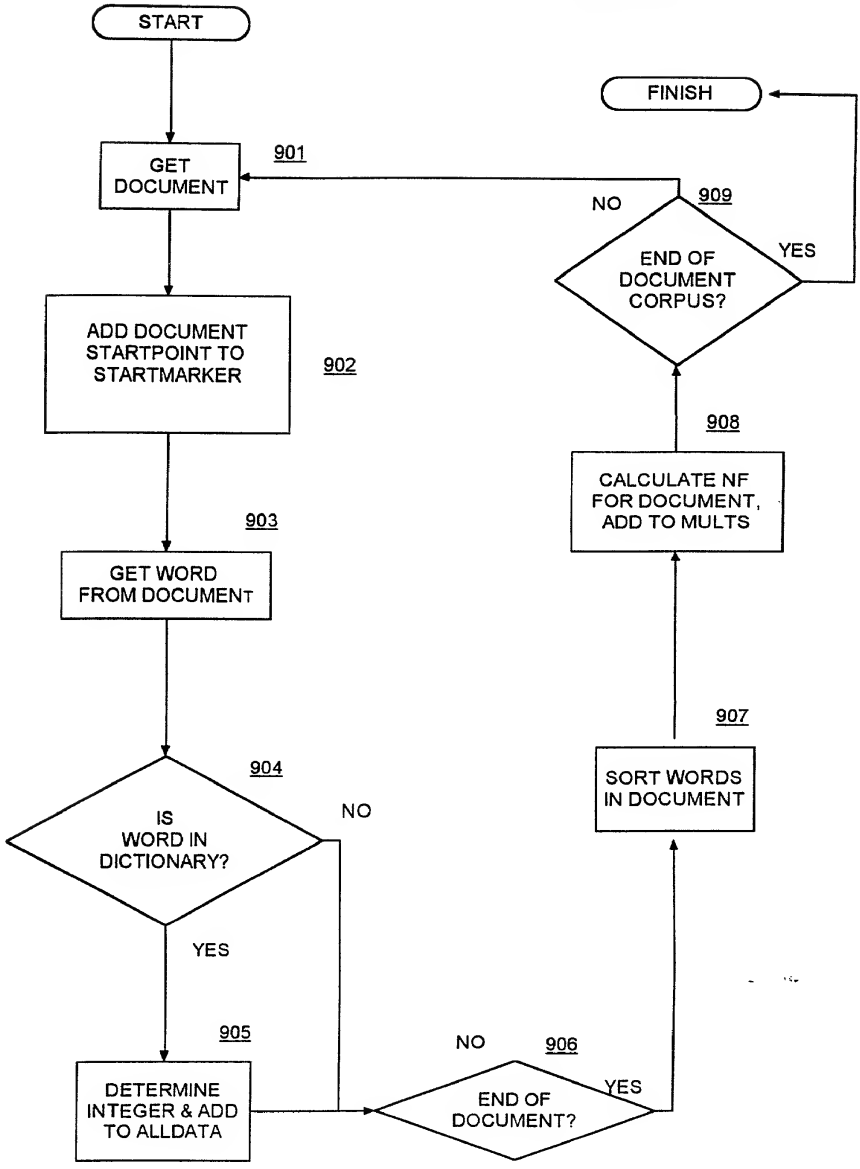
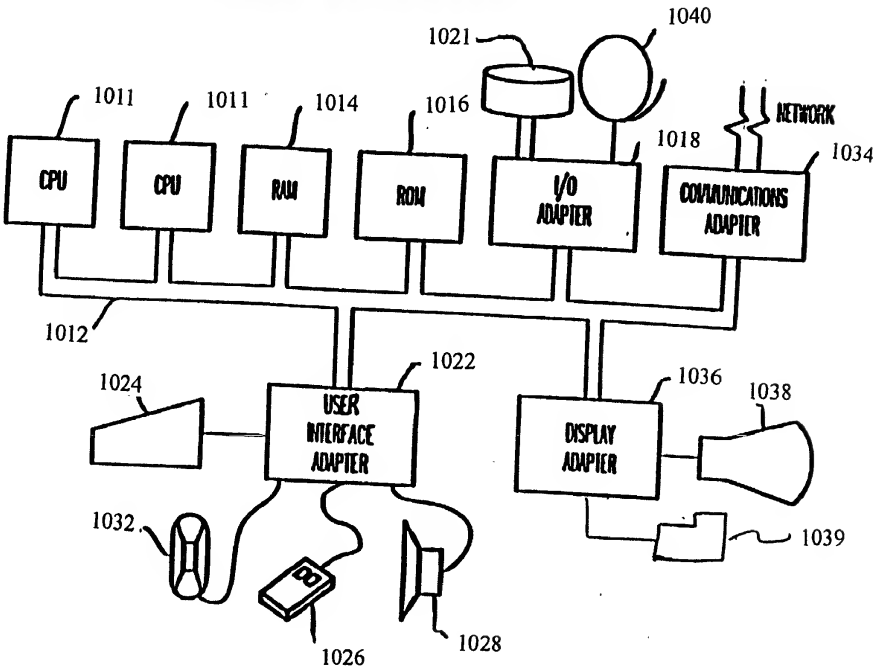

START
GET DOCUMENT
901
ADD DOCUMENT STARTPOINT TO STARTMARKER
902
GET WORD FROM DOCUMENT
903
IS WORD IN DICTIONARY?
904
NO
YES
DETERMINE INTEGER & ADD TO ALLDATA
905
END OF DOCUMENT?
906
NO
YES
SORT WORDS IN DOCUMENT
907
CALCULATE NF FOR DOCUMENT, ADD TO MULTS
908
END OF DOCUMENT CORPUS?
909
NO
YES
FINISH

FIGURE 10



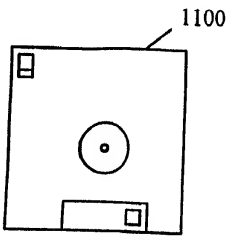


FIGURE 11